



FEBRUARY 2015

# GAAP Research Brief: Key Findings and Implications

The contents of this document were developed under an Enhanced Assessment Grant (#S368A120006) from the U.S. Department of Education. However, the contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal government.



It's all about student learning. **Period.**

100 Education Way  
P.O. Box 1217  
Dover, New Hampshire  
800.431.8901

©2015 Measured Progress. All rights reserved. Measured Progress and its logo are registered trademarks of Measured Progress, Inc. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as: Higgins, J., Famularo, L., Bowman, T. & Hall, R. (2015). *GAAP Research Brief: Key Findings and Implications* Retrieved [date] from <http://gaap.measuredprogress.org/gaap/>

# Table of Contents

Background and Purpose .....	4
Overview of GAAP .....	5
Research Methods .....	5
ASL Representation of Test Content: Key Findings .....	6
Randomized Controlled Trial Findings .....	6
Cognitive Lab Findings .....	6
Audio Representation of Test Content: Key Findings .....	7
Randomized Controlled Trial Findings .....	7
Cognitive Lab Findings .....	7
Implications .....	8
Approach for Future Guideline Development .....	8
ASL Item Development Process .....	8
Implications of Research Findings .....	8
Next Steps .....	11
References .....	12

## Background and Purpose

In the United States, more than 3.4 million students in tested grades (3 through 8 and high school) have Individual Education Plans (IEPs) and are served under the Individuals with Disabilities Education Act Part B (U.S. Department of Education, Office of Special Education Programs, 2013). While figures vary based on grade level, in 2010-11, between 37-46% of these students took their state's mathematics and language arts assessment based on grade level achievement standards with an accommodation; about 9% of the students took an alternate assessment (U.S. Department of Education, Office of Special Education Programs, 2013). That same year (2010-11), nearly 300,000 U.S. students ages 6 through 21 who were served under IDEA Part B were classified as hearing impaired or visually impaired, and a far greater number were classified as having a specific learning disability or multiple disabilities which may include blindness, deafness, hard of hearing, visual disabilities or print disabilities (U.S. Department of Education, Office of Special Education Programs, 2013).

Without access to appropriate accommodations and accessibility supports, students with vision impairments<sup>1</sup> (henceforth students with vision needs), students with reading/print disabilities (henceforth students with reading/print needs)<sup>2</sup>, English Learners (ELs), and Deaf and Hard of Hearing (DHH) students who communicate in American Sign Language (ASL) are placed at a disadvantage in demonstrating their academic proficiency on standardized tests. Historically, one of the main issues associated with providing audio and sign supports to students is the variability in administration by teachers or interpreters. This test administration variability can be attributed to teachers and interpreters not having ample time

to review the test to ensure they are familiar with the content, intentional or unintentional cueing of answers, variability in how content can be represented (e.g. mathematical notation or images), and variability in interpreters' ASL fluency and ability to translate English text into ASL. Today, most states have written guidelines regarding the roles and responsibilities of educators and other individuals who assist in the administration of standardized assessments delivered

“ Currently, the GAAP guidelines are the most comprehensive and only evidence-based guidelines in the field for representing K-12 test content in audio and ASL form. ”

via paper and pencil. In most cases, these guidelines fall short of the level of detail required to provide valid, fair, and equitable support for students who are better able to show what they know and can do when using supports. As states transition to digitally delivered assessments, accessibility supports that were previously delivered in person can be developed a priori and delivered digitally, thus eliminating many of these issues. However, the success of this approach is dependent on the ability of test developers to appropriately represent test content in different forms. At the time the Guidelines for Accessible Assessment Project (GAAP) was proposed, there were no comprehensive or research-based guidelines to aid test developers in creating audio and ASL representations of test content. The purpose of GAAP was to help fill this void.

Currently, the GAAP guidelines are the most comprehensive and only evidence-based guidelines in the field for representing K-12 test content in audio and ASL form. The GAAP guidelines are intended to aid state departments of education, assessment consortia, and test vendors in creating standardized audio and ASL versions of test items that provide quality access to students, while still measuring the same construct as is measured in the English text version of the assessment.

<sup>1</sup> Vision impairment means an impairment in vision that, even with correction, adversely affects a child's educational performance. The term includes both partial sight and blindness. (U.S. Department of Education, Sec. 300.8 Child with a disability, downloaded 1/5/15 at: <http://idea.ed.gov/explore/view/p/%2Croot%2Cregs%2C300%2CA%2C300%252E8%2C>)

<sup>2</sup> Print/reading disability describes persons certified by competent authority as unable to read or unable to use standard printed material as a result of physical limitations or as having a reading disability resulting from organic dysfunction and of sufficient severity to prevent their reading printed material in a normal manner (OSEP, NIMAS, and AIM Regulation Summary, 2013).

## Overview of GAAP

Over a period of two years, GAAP researchers engaged with partners from 18 state departments of education and expert advisory/working groups in the iterative development of evidence- and consensus-based guidelines and corresponding exemplar audio and ASL representations of test items. GAAP included a review of literature and current state practices on ASL and audio support in educational settings; expert review of draft guidelines and example ASL and audio representations of test items created based on the guidelines; cognitive lab and randomized controlled trial (RCT) research. All work products created through the GAAP grant are available on the [project website](#). The work products include: mathematics audio guidelines, ELA audio guidelines, English-text version of the ASL guidelines, ASL video version of the ASL guidelines, and example items with embedded audio and ASL support.

## Research Methods

Cognitive lab research was conducted with students who normally receive ASL or audio support during assessment to explore the impact of different ways of representing test content in audio and ASL form. The RCTs investigated the effect of providing computer-embedded ASL and audio support to students during testing (the intervention) on their performance on an assessment (the outcome). RCTs were conducted with students who normally receive audio or ASL support for assessment and a comparison group of students with no identified access needs (i.e. general education students). A stratified random sample design was employed for both the ASL and audio RCT. For the ASL RCT, teacher rating of DHH students' mathematics and reading ability formed the strata. For the audio RCT, student's audio access need formed the strata. Students within each stratum were randomly assigned to one of three test forms (see Table 1). Each participating student received the intervention (i.e. test items with ASL or audio support) and the control condition (i.e. test items with no support). Students in grades 3 through 12 from 26 states participated in GAAP cognitive lab and RCT research.

**Table 1. GAAP RCT Test Form Design**

Item Number	Form 1	Form 2	Form 3
1 – 6	Support variation 1	Support variation 2	Unsupported
7 – 12	Support variation 2	Unsupported	Support variation 1
13 - 18	Unsupported	Support variation 1	Support variation 2

Note: The ASL RCT included a fourth form for general education students that consisted of the same 18 items, in the same order but none of the items included ASL support.

# ASL Representation of Test Content: Key Research Findings

## Randomized Controlled Trial Findings

**General education students outperformed DHH students who normally use ASL support.** Students' scores on the 12 items delivered with support were used for the comparison. On each test form, general education students' mean score (Form 1 M=6.72, SD=2.82; Form 2 M=6.96, SD=3.06; Form 3 M=6.42, SD=3.02) was significantly higher ( $p<.001$ ) than DHH students' mean score (Form 1 M=3.59, SD=1.96; Form 2 M=3.22, SD=2.06; Form 3 M=2.55, SD=1.87).

**In aggregate, students across the three grade level bands, who normally receive ASL support performed better on the supported items than on the unsupported items.** Students' scores on the 12 items administered with support were compared to their scores on the 6 items administered unsupported (weighted for this analysis for a total possible score of 12). DHH students' mean score on the supported items (M=3.24, SD=2.04) was significantly higher ( $p=.013$ ) than on the unsupported items (M=2.91, SD=2.28,  $t(278)=2.24$ ,  $p=.013$ ).

## Cognitive Lab Findings

**DHH students who use ASL support prefer items where non-textual content such as equations, expressions, and graphics are signed over items where non-textual content is not signed.** When asked to express a preference for how non-textual content should be presented, the majority of students across all three grade level bands reported preferring ASL representation of this information. Students reported that items with non-textual content signed are more consistent with how this information is presented during instruction, and consistent with ASL as a visual language and how ASL is used in the deaf community.

**DHH students who use ASL support prefer items using a diamond structure.** A common way of structuring test items is to present several sentences or a table of information followed by a question. Items set up using the diamond format introduce the question or goal first and restate it at the end, with item information in between. DHH students reported a preference for items signed in diamond structure, stating that the format was more like ASL than English. Students found items using the non-diamond structure confusing and more difficult to understand than items with the diamond structure.

**DHH students who use ASL support prefer mathematical terms to be either signed or signed and fingerspelled, as opposed to just fingerspelled.** In cases when students knew the ASL sign for a particular term, they preferred the term being signed rather than signed and fingerspelled. Students who may not have known the sign for a particular term preferred the combination of sign and fingerspelling. In particular, students at lower grade levels responded positively to fingerspelling with signing. The fingerspelling only condition was studied at the high school level and none of the 15 participating students reported a preference for mathematical terms being fingerpelled only.

## Audio Representation of Test Content: Key Findings

### Randomized Controlled Trial Findings

**General education students outperformed students who normally use audio support.** Students' scores on the 12 items delivered with support were used for the comparison. General education students mean score ( $M=7.09$ ,  $SD=2.91$ ) was significantly higher ( $p<.001$ ) than audio needs students' mean score ( $M=4.97$ ,  $SD=2.71$ ).

**In aggregate, students across the three grade level bands, who normally receive audio support performed better on the supported items than on the unsupported items.** Students' scores on the 12 items administered with support were compared to the scores on the 6 items administered unsupported (weighted for this analysis for a total possible score of 12). Students' mean score on the supported items ( $M=5.00$ ,  $SD=2.71$ ) was significantly higher ( $p=.025$ ) than on the unsupported items ( $M=4.25$ ,  $SD=3.07$ ,  $t(64)=2.01$ ,  $p=.025$ ).

### Cognitive Lab Findings

**Students with audio access needs (vision need, reading/print need, English Learner), reported a preference for having mathematical notation such as expressions, equations, and numerals read aloud versus not read.** Overwhelmingly, students with vision needs preferred to have mathematical notation read. Students with reading/print needs and English Learners were mixed in their preferences for having expressions, equations, and numerals read aloud, but overall the majority of students reported a preference for having mathematical notation read. Some students indicated that when expressions, equations, and numerals were read aloud, the item flowed better and items where they were not read aloud were difficult to follow.

**Students' preferences for audio representations of graphics and images differed based on their access needs.** Students in each of the three groups (vision need, reading/print need, English Learner) have unique reasons for needing audio representation to better access test content. A middle school student with a vision need explained that when information on a graph is not read aloud, "my face is close to the screen and [I] lose numbers." A high school student explained that the image description provided similar information to what a sighted student would receive by looking at the item: The audio "described every detail in the picture instead of having to look at it, picture it in your head. People with low vision could see what is in the picture this way." However, some students with reading/print needs found the detailed descriptions of graphs and image descriptions to be cumbersome or confusing.

## Implications

In addition to the guidelines themselves, GAAP made several other important contributions to the field of accessible assessment. First, lessons learned during the development of the GAAP guidelines have enabled the GAAP team to recommend an effective approach for future guideline development. Second, the GAAP team established an approach for developing high quality ASL representations of test content based on lessons learned over the two years that the GAAP team members worked together, including the team thinking critically about and discussing the process. Third, GAAP makes a unique contribution to the assessment accessibility research base.

### Approach for Future Guideline Development

Building on learnings from a literature review, review of state guideline documents, and input from GAAP state partners and experts in the GAAP advisory/working groups, GAAP researchers developed a process for creating consensus-based audio and ASL guidelines that takes findings from research and best practices into consideration. The key factor in the process is engaging a multi-disciplinary team consisting of individuals with expertise in accessibility, assessment, content, and the particular access support or accommodation and population of students who use them, and facilitating meetings where team members have the opportunity to present their view on support decisions. The multi-disciplinary team meetings provide a forum to discuss the test content in question, students' needs, common practices for providing the support, relevant research, and the pros and cons of different approaches to providing the support. During GAAP, both the ASL and audio teams used the guiding principal that when applied to test content, the guidelines must result in items that are correct from a content standpoint, are linguistically correct, and provide the best access to students who use audio and ASL supports.

### ASL Item Development Process

The GAAP team learned many lessons through the process of developing high quality ASL representations of test items. The first is the importance of engaging with qualified experts to form a multi-disciplinary

team. Creating high quality ASL interpretations of test items requires collective expertise in multiple areas, including ASL, the grade level and content area being assessed, educational assessment and measurement, and quality video production. The team should include a deaf content expert who is also an educator with native ASL fluency, a certified bilingual interpreter with expertise in interpreting for educational assessment, a content specialist who has teaching and item development experience, and an assessment and accessibility specialist, who has training and experience in educational assessment and measurement and a thorough understanding of issues related to test reliability and validity for students using access and accommodation supports. The team should also include a video production specialist who has experience in filming ASL. Throughout the process, open communication, respect for the expertise that each team member brings to the table, positive team dynamics and consensus decision-making are critical.

A second lesson learned is the importance of using a thoughtful process in developing ASL interpretations of the test items to ensure high quality representations. This team-based collaborative process should include a review of English test items and discussion of potentially controversial interpretation issues; development and review of draft ASL representations of items identified as potentially challenging or controversial; recording of ASL videos for all items; review of videos by external ASL and content experts. Ongoing research and evaluation is also recommended in order to make improvements to the ASL Guidelines and subsequent ASL videos of assessment content. For a detailed description of the ASL item development team and process recommended by the GAAP team, see the [Guidelines for the Development of American Sign Language Versions of Academic Test Content for K-12 Students](#).

### Implications of Research Findings

Before discussing the implications of the GAAP research findings, it is important to point out that GAAP research makes a unique contribution to the field, in part, because it differs from previously conducted research on audio and ASL supports in several important ways. These differences, which likely impacted the findings, relate to the quality of the audio



and ASL item representations used for this research. First, unlike other audio and ASL support research, the overarching goal of GAAP was to create a set of audio and ASL guidelines that, when applied to test content, would provide high quality access while not altering the construct measured by the item. The initial set of guidelines were informed by a literature review, best practices gleaned from a review of existing state guidelines, and decisions made during consensus-driven discussions among a multidisciplinary team of experts. The research was designed to further inform guideline development by examining the impact of different guidelines for representing the same content, and the extent to which the application of the guidelines to test items creates representations that remove construct irrelevant barriers to student performance. Second, the audio and ASL item representations used for GAAP research were created by a multidisciplinary team of experts with a focus on providing access while not altering the constructs measured by the item, leading or cueing students to a particular response, or giving an unfair advantage or disadvantage to the students who receive the supported version. And, in the case of ASL, the team adhered to the linguistic rules and conventions of ASL while not altering the constructs measured by the items. The multidisciplinary team was also physically present during the taping of the ASL videos to ensure quality. Third, GAAP research was conducted using a computer-based testing system with embedded audio and ASL support. Unlike in-person ASL or audio administration, this method ensures that students are delivered the same representations of the items. In addition, the system enabled students to read the English test item, turn the support on or off, and replay the entire audio or ASL item or portions of it as many times as they like. In other words, each student was able to decide how to access the test content in a way that best fit his or her needs. Each of these contributed to high quality audio and ASL representations of the test items used for both the cognitive lab and RCT research.

In addition to the differences described above, GAAP research findings make several unique and important contributions to the research base. Our literature review revealed few published studies on the impact of the ASL support and these studies have shown no significant difference in DHH students' performance on supported and unsupported items

(Cawthon, Winton, Garberoglio, & Gobble, 2011). These researchers hypothesize the reason for lack of evidence of improved scores with ASL support may be related to DHH students' problem solving ability (Ansell & Pagliaro, 2006), the adequacy of the ASL translation, whether the form of ASL used in the test differs from conversational and instructional ASL (Cawthon, Winton, Garberoglio, & Gobble, 2011), and how students actually utilized ASL support during the assessment (Cawthon, Winton, Garberoglio, & Gobble, 2011).

“...the audio and ASL item representations used for GAAP research were created by a multidisciplinary team of experts with a focus on providing access while not altering the constructs measured by the item, leading or cueing students to a particular response, or giving an unfair advantage or disadvantage to the students who receive the supported version.”

To the best of the researchers' knowledge, the GAAP ASL RCT is the first study to find a significant difference in DHH students' performance on supported and unsupported items. This finding suggests that providing high quality ASL support to DHH students who normally receive ASL support during assessment reduces construct irrelevant barriers and provides greater access to the test content. One advantage of GAAP over other research among this population is the GAAP study had a relatively large sample size. A total of 279 DHH students participated in the GAAP RCT, which is uncommonly large for this relatively small student population. This contributed to the statistical power of the study to detect a difference between students' performance on the supported and unsupported items. Second, the GAAP research was not impacted by some of the factors other researchers

have cited as possible reasons for finding no significant difference between student performance on supported and unsupported items (described above). In particular, the ASL translations were high quality and consistent with how ASL is used in instruction, and the ASL support was delivered consistently and in a manner that enabled students to decide when to utilize it.

Key findings from the ASL cognitive lab research shed light on issues related to translating test content into ASL. Whenever English test items are translated into another language, it raises concerns about whether the translation will alter the meaning of the item, lead or cue students to a particular response, or otherwise give an unfair advantage to the students who receive the translated version. In order for test scores on translated tests to be comparable with those from assessments administered in English, it is imperative that the translated test items represent the content presented in the English text in a way that does not change what is being measured. Maintaining the meaning of the test item does not entail a literal or direct word for word translation of the English text into the other language; in fact, this is highly likely to alter the original meaning. Findings from the cognitive labs support the idea that the translation should adhere to the linguistic rules and conventions of the language into which the items are being translated. Specifically, findings from the cognitive labs suggest that DHH students are better able to understand items that 1) use ASL conventions related to the order in which information is presented (e.g. the diamond set up), 2) are consistent with how ASL is used during instruction (e.g. non-textual content such as mathematical notation and graphs is signed), and 3) are consistent with ASL conventions related to the use of fingerspelling.

The findings from the audio studies contribute to the base of research on the validity of inferences made from supported items for students with vision needs and print/reading needs, and ELs. Various methods (e.g. differential boost, measurement comparability, and item-level effects) have been used to study read aloud, with mixed findings about the effectiveness in removing construct-irrelevant variance associated with reading difficulties on mathematics tests (Bolt & Thurlow, 2006; Rogers, Christian & Thurlow, 2012). Our analysis of audio RCT data revealed that students who normally receive audio support perform better on items that contained audio support than on items that did not contain audio support. This finding suggests

that providing high quality audio support to students who normally receive it during assessment reduces construct irrelevant barriers and provides greater access to the test content.

Key findings from the audio cognitive lab research have significant implications for the delivery of audio support through computer-based testing systems. First, across grade levels and category of need, students reported preferring full audio access to expressions and equations. The GAAP team hypothesized that full audio representations of expressions and equations could be overly cumbersome and was concerned that the audio support of this mathematics content might not be consistent with instruction however, during the cognitive labs students consistently reported preferring items where the expressions and equations were read compared to items where they were not read aloud. Conversely, the cognitive lab research revealed that for certain types of mathematics content, multiple versions of support are needed in order to provide high quality access to different groups of students. Specifically, detailed descriptions of graphics and images are required for students with vision needs, but are not necessary and might be distracting for students with print/reading needs and ELs. The need for differentiation in audio support based on student needs has two key implications for the digital delivery of audio support. First, states and consortia need to develop multiple versions of audio representation of test content and ensure that the test delivery system has the ability to provide students with the appropriate version, as predefined by a teacher or test administrator. Second, teachers and test administrators need training on how to determine which audio version should be provided to individual students based on their audio access needs. States and consortia should consider these important implementation and policy issues when defining their needs for audio support. The second key finding is that students in need of audio support expressed a desire for accessing the test content in a more personalized way that provides a high degree of independence. Some examples of personalized audio support are providing students the ability to adjust pacing, choose the type of voice (male or female, human or synthesized), and the ability to have the item automatically read upon entry to the item or have text read only when selected by the student.

## Next Steps

Currently, the GAAP guidelines are the most comprehensive and only evidence-based guidelines in the field for representing K-12 test content in audio and ASL form. While these guidelines are an important contribution to the field of accessible assessment, there is more guideline research and development work to be done. First, work is needed to develop comprehensive, evidence-based guidelines in additional content areas, such as science, which includes technical terms and scientific images, and social studies, which includes maps and timelines. Second, future audio guideline development should include additional users of audio support such as blind and hard-of-hearing students who were not included in GAAP due to time and budget limitations. Future audio guideline development research should include a braille research form and a new category of audio support for hard-of-hearing students. GAAP research findings underscore the importance of understanding the unique needs of different groups of students thus, it is critical that blind and hard-of-hearing students are included in future audio support research and guideline development. Third, future audio guideline work should take into account that states and consortia are transitioning to a model where use of audio support is not dictated by students' disability status and conduct research with students who use audio support, but do not have an IEP to more deeply understand these students' access needs. Lastly, to enable all students with access needs to fully participate in next generation assessments and to increase the validity of test score-based inferences about their academic proficiency, more research and development is needed to create guidelines for additional accessibility supports and item representations including but not limited to tactile overlays and visual representations of test content.

GAAP also revealed the need for additional ASL support research and development. First, more work is needed to develop best practices in ASL support delivery. Specifications such as size of the video window and features such as bookmarking of video segments should be researched and documented. Second, an ASL glossary for assessment terminology to ensure standardized ASL delivery of test content

is needed. To date, there is not a widely accepted glossary for content terminology (e.g. quadrilateral) or assessment terminology (e.g. drag and drop) for which there are no widely recognized ASL signs. A multidisciplinary team of assessment, accessibility, content, and ASL experts is needed to collect key terms from states and consortia, and develop a research-based glossary of terms that can be both used for test development purposes and disseminated to students and teachers for use during instruction.

“ ... while GAAP added evidence to the field of assessment on the positive impact of providing high quality and individualized computer-based audio and ASL support to students, more work is needed to level the playing field for students in need of assessment and instructional accessibility supports. ”

Lastly, while GAAP added evidence to the field of assessment on the positive impact of providing high quality and individualized computer-based audio and ASL support to students, more work is needed to level the playing field for students in need of assessment and instructional accessibility supports. In both the audio and ASL RCT, general education students significantly outperformed students with identified audio and ASL access needs with DHH students answering 26% of supported items correctly, students in need of audio support answering 41% of supported items correctly, and general education students answering 57% of items correctly. This disparity is not new (see Cawthon, Winton, Garberoglio, and Gobble, 2011 and Johnstone, et al, 2012) and highlights the need to focus attention on both assessment and instructional needs of DHH students and students who use audio support to access content.

## References

- Ansell, E., & Pagliaro, C. M. (2006). The relative difficulty of signed arithmetic story problems for primary level deaf and hard-of-hearing students. *Journal of Deaf studies and Deaf Education*, 11(2), 153-170.
- Bolt, S. E., & Thurlow, M. L. (2006). *Item-level effects of the read-aloud accommodation for students with reading disabilities (Synthesis Report 65)*. Minneapolis, MN: University of Minnesota.
- Cawthon, S, Winton, S., Garberoglio, C. L., Gobble, M. (2011). The Effects of American Sign Language as an Assessment Accommodation for Students Who Are Deaf or Hard of Hearing. *Journal of Deaf Studies and Deaf Education*, 16(2), 199-211.
- Johnstone, C., Rogers, C., Wu, Y.-C., Fedorchak, G., Katz, M., & Higgins, J. (2012). *Rules for audio representation of science items on a statewide assessment: Results of a comparative study (Technical Report 64)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Office of Special Education Programs, NIMAS, and AIM Regulation Summary, 2013, retrieved on 1/7/15 from: [http://aim.cast.org/learn/policy/federal/osep\\_summary#foot2](http://aim.cast.org/learn/policy/federal/osep_summary#foot2)
- Rogers, C. M., Christian, E., & Thurlow, M. L. (2012). *A summary of the research on the effects of test accommodations: 2009-2010 (Technical Report 65)*. Minneapolis MN: University of Minnesota, National Center on Educational Outcomes.
- U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, *35th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2013*, Washington, D.C. 2014.
- Weigert, S. (2012). Aligning and Inventing Practices to Achieve Inclusive Assessment Policies: A decade of work toward optimal access for US students with disabilities 2001-2011. *International Journal of Disability, Development, and Education*, 59, 1, 21-36.